

Automated Grape Classification Using a YOLOv11n-Based Computer Vision System

Marcela Siomara Recuento Tapia , Bryan Alexandre Salazar Pineda , Ryan Abraham León León 
Universidad Privada del Norte, Perú, n00303611@upn.pe, n00141117@upn.pe, ryan.leon@upn.edu.pe

Abstract– This study presents the implementation and evaluation of a computer vision system based on the YOLOv11n model for the automated classification of grapes intended for wine production at a Peruvian agroindustrial company. The proposed system replaces manual visual inspection, whose average precision is approximately 76%, with a lightweight real-time detection model trained on 5,137 augmented images annotated using Roboflow. The model achieved a mean precision of 89.1% (95% CI: 87.4–90.8%), a recall of 88.3% (95% CI: 86.1–90.4%), and an mAP@50 of 92.5% (95% CI: 91.2–93.8%) for the Premium grape category, representing a 13% improvement in precision compared to traditional manual inspection. Lower performance was observed for the Bulk category (76.1% precision and 61.4% recall; 95% CI: 58.0–64.8%), mainly due to morphological variability among the grape clusters. The results confirm the technical feasibility and scalability of YOLOv11n as an effective tool for optimizing quality control processes in viticulture, demonstrating its capability for real-time inference under uncontrolled environmental conditions.

Keywords-- Artificial vision, YOLOv11n, grape classification, agricultural automation, quality control, deep learning.

Clasificación automatizada de uvas mediante un sistema de visión artificial basado en YOLOv11n

Marcela Siomara Recuento Tapia[✉], Bryan Alexandre Salazar Pineda[✉], Ryan Abraham León León[✉]
Universidad Privada del Norte, Perú, n00303611@upn.pe, n00141117@upn.pe, ryan.leon@upn.edu.pe

Resumen— Este estudio presenta la implementación y evaluación de un sistema de visión artificial basado en YOLOv11n para la clasificación automatizada de uvas destinadas a la producción de vinos en una empresa agroindustrial peruana. El sistema sustituye la inspección visual manual, cuyo nivel de acierto promedio es del 76%, mediante un modelo ligero de detección en tiempo real entrenado con 5,137 imágenes aumentadas y etiquetadas en Roboflow. El modelo alcanzó una precisión promedio del 89.1% (IC95%: 87.4–90.8%), un recall del 88.3% (IC95%: 86.1–90.4%) y un mAP@50 del 92.5% (IC95%: 91.2–93.8%) para las uvas “Premium”, lo que supone una mejora del 13% en precisión respecto al método manual tradicional. En la clase “Granel” se obtuvo un rendimiento inferior (76.1% de precisión, 61.4% de recall, IC95%: 58.0–64.8%) debido a la variabilidad morfológica. Los resultados confirman la viabilidad y escalabilidad del modelo YOLOv11n como herramienta para optimizar el control de calidad en viticultura, con capacidad de inferencia en tiempo real bajo condiciones no controladas.

Palabras clave-- visión artificial, YOLOv11n, clasificación de uvas, automatización agrícola, control de calidad, aprendizaje profundo.

I. INTRODUCCIÓN

La clasificación de uvas es una etapa fundamental dentro de los procesos vitivinícolas, ya que determina la calidad de la materia prima destinada a la elaboración de vinos y otros derivados [1]. Tradicionalmente, esta actividad se realiza mediante inspección visual manual, un método dependiente de la experiencia del operario y sujeto a variabilidad por condiciones de iluminación, subjetividad en los criterios de evaluación, fatiga y limitaciones de tiempo, lo que puede comprometer la consistencia de los estándares de calidad y dificultar la escalabilidad del proceso cuando se manejan grandes volúmenes de fruta [2][3]. En los últimos años, la visión artificial ha adquirido un rol prominente en la automatización de tareas agrícolas, permitiendo analizar características como color, textura, forma y defectos superficiales mediante técnicas de procesamiento digital de imágenes y modelos de aprendizaje profundo [4]. Diversos estudios han demostrado que las redes neuronales convolucionales ofrecen un desempeño superior frente a métodos tradicionales, ya que proporcionan mayor robustez ante variaciones en las variaciones ambientales y permiten operar con altos niveles de precisión. Sin embargo, la mayoría de las investigaciones se centran en la detección general de frutas o en el análisis fenotípico de racimos completos, sin profundizar en la clasificación específica de uvas según categorías de calidad comercial, especialmente bajo escenarios reales con oclusiones, fondos irregulares y cambios abruptos de

iluminación [5] [6][7]. Dentro de los avances más destacados en visión artificial se encuentran los modelos de la familia YOLO (You Only Look Once), y cuya versión más reciente, YOLOv11, junto con su variante ligera YOLOv11n, ha ampliado la posibilidad de implementar sistemas eficientes en hardware de bajo costo, aspecto particularmente relevante para la agroindustria [10]. No obstante, la aplicación de estos modelos para la clasificación detallada de uvas destinadas a procesos vitivinícolas continúa siendo limitada y representa una línea de investigación con amplias oportunidades de desarrollo [11][12]. En este contexto, el presente trabajo se centra en el desarrollo de un sistema de visión artificial basado en YOLOv11n, orientado específicamente a la clasificación automática de uvas en dos categorías representativas de la industria vitivinícola: Premium y Granel, bajo un enfoque metodológico y técnico, sin asociarlo a la resolución de un problema empresarial directo. Se detalla el proceso de construcción del dataset, la metodología de etiquetado, la estrategia de aumento de datos, la configuración del entrenamiento del modelo y la evaluación de su desempeño bajo distintas condiciones reales de captura, como variabilidad de iluminación, ángulos de observación y niveles de oclusión [13][14]. El objetivo principal de esta investigación es demostrar la viabilidad técnica del uso de YOLOv11n para la clasificación automatizada de uvas y documentar un pipeline reproducible que pueda servir como base para futuras implementaciones en ambientes de producción, analizando además las principales fuentes de error identificadas durante la validación y posibles estrategias de mejora, tales como minería de errores, ajustes de umbral por clase y técnicas de aprendizaje sensible al costo. Finalmente, se discuten líneas de trabajo futuras orientadas a la integración del sistema en escenarios operativos reales, incluyendo la recolección dirigida de nuevos datos, pruebas piloto en planta y el monitoreo de la deriva del modelo en campañas agrícolas posteriores, aportando a la literatura científica un desarrollo metodológico claro, replicable y basado en modelos de última generación, y demostrando que herramientas ligeras como YOLOv11n pueden adaptarse eficazmente a tareas agrícolas especializadas y operar con precisión en condiciones no controladas.

II. MATERIALES Y MÉTODOS

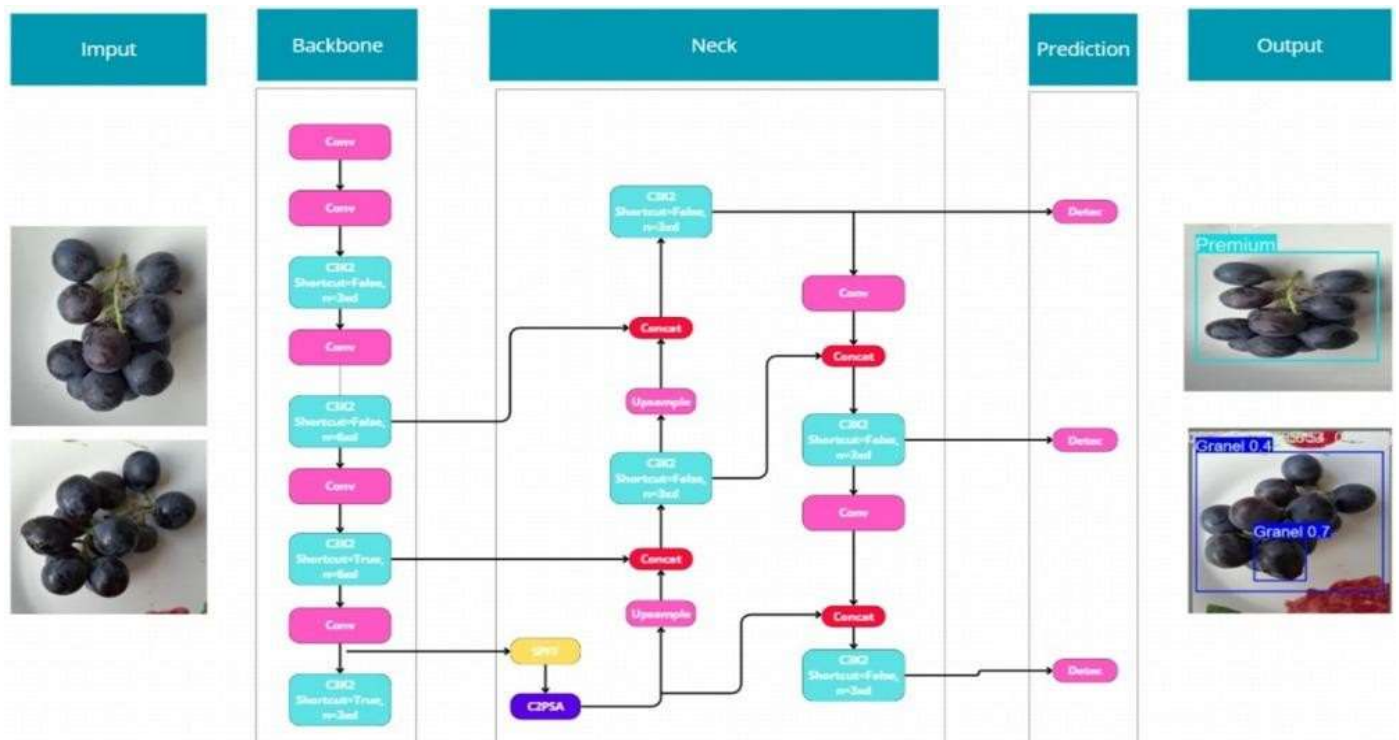
El presente estudio tuvo como objetivo implementar y validar un sistema de visión artificial basado en YOLOv11n. Se eligió esta versión en particular debido a que ofrece un equilibrio óptimo entre precisión y velocidad de inferencia, aspectos críticos en aplicaciones de tiempo real. A diferencia de

Lactecy T4 Tensar RT10 PP16 (mg/kg)	YOLO11 COCO and %w/w	YOLO10 COCO and %w/w	YOLOv11 COCO and %w/w	YOLOv12 COCO and %w/w	YOLOv13 COCO and %w/w	YOLOv14 COCO and %w/w	YOLOv15 COCO and %w/w	PP10v16 COCO and %w/w
0	38	38	38	38	38	38	38	38
2	48	48	48	48	48	48	48	48
4	58	58	58	58	58	58	58	58
6	68	68	68	68	68	68	68	68
8	78	78	78	78	78	78	78	78
10	88	88	88	88	88	88	88	88
12	94	92	88	86	84	82	80	78
14	94	92	88	86	84	82	80	78
16	94	92	88	86	84	82	80	78
18	94	92	88	86	84	82	80	78

Se recopilaron un total de 2,125 imágenes originales de racimos de uvas en condiciones reales. Las imágenes fueron capturadas con un iPhone 13 Pro Max, equipado con una cámara principal de 12 MP (f/1.5, 26 mm, sensor 1/1.65"),

Las imágenes se capturaron desde tres perspectivas principales para asegurar variabilidad en el dataset: una perspectiva cenital a 90°, una lateral a 45° y una oblicua a aproximadamente 30°. La perspectiva cenital permitió obtener una vista directa del racimo y facilitar la identificación completa de su estructura; la vista lateral aportó información sobre el volumen y la disposición tridimensional de las bayas; y la vista oblicua simuló condiciones típicas de operación en una línea de selección, donde las cámaras no siempre se encuentran perfectamente alineadas respecto al objeto. La combinación de estos tres ángulos permitió capturar un conjunto de imágenes más diverso y representativo de escenarios reales.

Esta estrategia garantizó la independencia entre conjuntos y evitó el sesgo temporal asociado a la variabilidad lumínica o de exposición inherente al dispositivo móvil.



3



Fig. 3 Racimos de uvas capturados en campo bajo luz natural desde distintos ángulos para alimentar el modelo de clasificación.

La clasificación utilizada para el etiquetado incluyó únicamente dos categorías: Premium y Granel. La clase Premium (fig. 4), comprendió racimos íntegros, sin defectos visibles, con color uniforme y sin presencia de hojas o tallos adheridos. La clase Granel (fig. 5) representó racimos con daños físicos, variaciones de color, presencia de hojas, tallos o cualquier característica que redujera su calidad comercial. La anotación de las imágenes se realizó en formato YOLO, empleando coordenadas normalizadas para la ubicación de los bounding boxes. Esta clasificación se estableció con base en revisiones conjuntas entre los anotadores, logrando un alto nivel de consistencia en la interpretación de los criterios de calidad.



Fig. 4 Etiquetado de uvas premium.

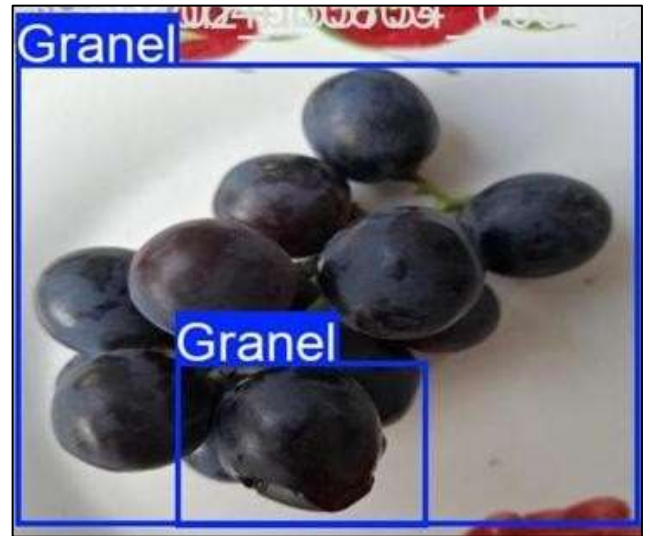


Fig. 5 Etiquetado de uvas granel.

Cada imagen contenía entre 1 y 5 racimos anotados mediante cuadros delimitadores (bounding boxes) en formato YOLO (x_center, y_center, width, height, class_id), revisados por consenso entre ambos anotadores. Adicionalmente, se priorizó un ajuste estricto de los cuadros al contorno del racimo para minimizar el ruido de fondo y se establecieron criterios para filtrar oclusiones severas, asegurando que cada instancia etiquetada poseyera visibilidad suficiente para la correcta extracción de características. La consistencia inter-anotador se verificó sobre el 10 % del dataset, obteniendo un índice de concordancia Cohen's $\kappa = 0.92$, considerado excelente.

Para mejorar la generalización del modelo, se aplicaron transformaciones de data argumentation en Roboflow con los parámetros y probabilidades mostrados en la Tabla 1. Estas transformaciones se aplicaron online durante el entrenamiento, preservando la proporción original de clases.

Tabla 1
Transformaciones de aumento de datos aplicadas en Roboflow

Transformación	Rangos	Probabilidad
Flip horizontal / vertical	-	0.50
Crop aleatorio	Zoom 0 % - 29 %	0.40
Shear	$\pm 15^\circ$ horizontal / $\pm 15^\circ$ x x vertical	0.30
Escala de grises	22% de las imágenes	0.40
Saturación	-30% a + 30 %	0.40
Desenfocado gaussiano	≤ 3.3 px	0.25
Brillo / contraste	± 20 %	0.30
Rotación	$\pm 10^\circ$	0.30

La división del dataset se realizó de manera estratificada para mantener la proporción de clases en cada subconjunto. El conjunto se separó en entrenamiento, validación y prueba,

garantizando que todas las imágenes pertenecientes a un mismo racimo permanecieran dentro del mismo subconjunto para evitar filtraciones entre fases. Esta separación se efectuó utilizando una semilla fija, lo que permite replicar la misma distribución en futuras ejecuciones del proceso. La proporción elegida respondió a la necesidad de disponer de un conjunto amplio para el entrenamiento, un subconjunto equilibrado para la validación y un conjunto independiente destinado exclusivamente a la evaluación final.

Las especificaciones técnicas del hardware y software utilizados se detallan en la Tabla 2.

Se empleó estratificación por clase y aleatorización controlada (seed = 42) para reproducibilidad. Cada subconjunto mantuvo la proporción aproximada de 70 % Premium y 30 % Granel.

Tabla 2
Especificaciones del entorno de entrenamiento

Componente	Descripción
CPU	Intel Core i5-10400F @ 2.9 GHz
GPU	NVIDIA RTX 4060 Ti (8 GB VRAM)
RAM	16 GB DDR4
SO	Windows 11 Pro-64-bit
Framework	PyTorch 2.7.0
Implementación YOLO	Ultralytics v8.3.137 (YOLOv11n branch)
Phyton	3.13.0
Librerías	OpenCV 4.10.0, NumPy 1.26.4, Matplotlib 3.9.0, Roboflow 1.1.0

El modelo YOLOv11n se eligió por su equilibrio entre precisión y velocidad de inferencia en hardware de consumo. La configuración detallada se presenta en la Tabla 3.

Tabla 3
Parámetros de entrenamiento del modelo YOLOv11n

Parámetro	Valor	Descripción
Tamaño de imagen	640 x 640 px	Resolución estándar de entrada
Batch size	16	Imágenes procesadas por iteración
Optimizador	SGD (momentum = 0.937)	Recomendado por Ultralytics
Épocas máximas / reales	100/30	Early stopping tras 10 épocas sin mejora
NMS threshold (IoU)	0.45	Eliminación de detecciones redundantes

Umbral de confianza	0.276	Valor óptimo obtenido del análisis F1
---------------------	-------	---------------------------------------

La arquitectura seleccionada fundamenta su operación en una red neuronal convolucional profunda, donde la extracción de características visuales de los racimos se realiza mediante la operación de convolución discreta. Matemáticamente, para una imagen de entrada I y un kernel K de dimensiones $m * n$, el mapa de características resultante S en la posición (i, j) se define mediante la ecuación:

$$S(i, j) = \sum_{u=0}^{m-1} \sum_{v=0}^{n-1} I(i+u, j+v) * K(u, v) \quad (1)$$

Esta operación permite identificar patrones esenciales como la textura de la piel y los bordes de las bayas, diferenciando así la clase Premium de la Granel. Dicho proceso se ilustra en la fig. 6, donde se observa la transformación de la imagen. Para gestionar la no linealidad de los datos y adaptarse a la variabilidad morfológica de las uvas, el modelo integra la función de activación SiLU (Sigmoid Linear Unit) en sus capas ocultas, cuya expresión se formula como:

$$f(x) = x * \sigma(x) = \frac{x}{1 - e^{-x}} \quad (2)$$

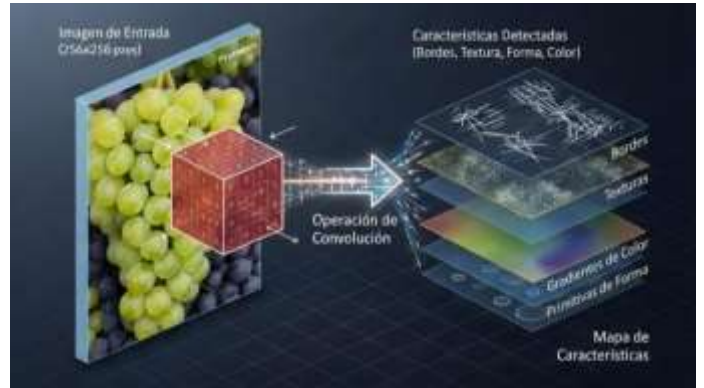


Fig. 6: Representación tridimensional de la operación de convolución aplicada a un racimo para la extracción de características visuales.

Esta función permite a la red aprender mapeos complejos y suaves, superiores a la activación ReLU tradicional en redes profundas. El proceso de aprendizaje fue guiado por la minimización de una función de pérdida compuesta (LTotal), que agrega ponderadamente la pérdida de localización (LBox), clasificación (Lcls) y detección (Ldfl):

$$L_{total} = \lambda_{box}L_{box} + \lambda_{cls}L_{cls} + \lambda_{dfl}L_{dfl} \quad (3)$$

Donde los coeficientes λ equilibran la importancia de ajustar los cuadros delimitadores y clasificar correctamente las categorías. Finalmente, la optimización de los pesos θ se ejecutó mediante el algoritmo SGD (Stochastic Gradient Descent, por sus siglas en inglés) con Momentum, regido por la regla de actualización:

$$v_t = \mu v_t + \nabla L(\theta_t), \theta_{t+1} = \theta_t - \alpha v_{t+1} \quad (4)$$

Donde α es la tasa de aprendizaje y μ es el factor de momento establecido en 0.9371, un parámetro crítico que permitió acelerar la convergencia y amortiguar las oscilaciones del gradiente ∇L durante las 30 épocas de entrenamiento efectivo. Adicionalmente, se aplicaron parámetros de posprocesamiento rigurosos, como un umbral de confianza de 0.276 y un umbral de Non-Maximum Suppression (NMS) con IoU (Intersection over Union) de 0.45, esenciales para eliminar detecciones redundantes en racimos densamente agrupados.

El rendimiento se evaluó con las métricas estándar: Precisión (P), Recall (R), F1-Score, y Mean Average Precision (mAP@50 y mAP@50:95). El comportamiento del modelo respecto a estas métricas se visualiza en la curva de Precisión-Recall (Fig. 7) y en la curva de F1-Confianza (Fig. 8)

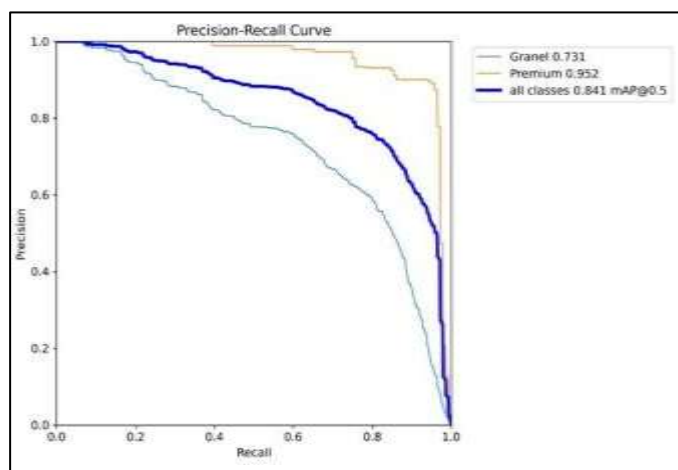


Fig. 7 Curva de precisión-recall

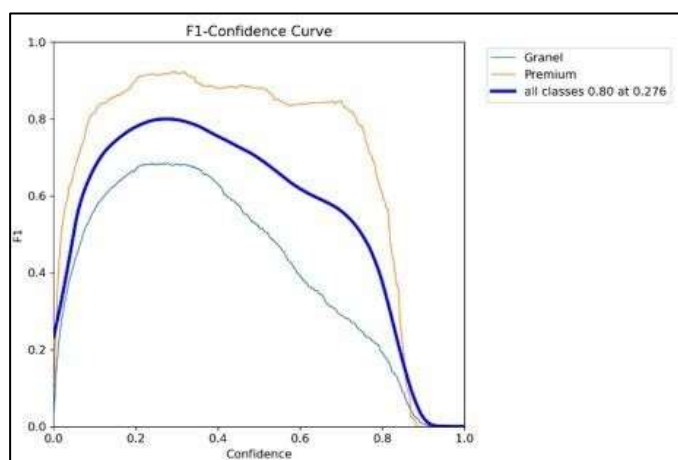


Fig. 8 Curva F1-confianza.

El umbral de confianza óptimo se determinó analizando las curvas F1-Confidence y Precision-Recall sobre el conjunto de validación. El máximo F1 se obtuvo en 0.276, adoptado como umbral operativo para balancear precisión y exhaustividad.

Dicho umbral se mantuvo fijo en todas las pruebas y evaluaciones.

El proceso de inferencia aplicó Non-Maximum Suppression (NMS) con IoU = 0.45 y ordenamiento descendente por probabilidad de clase, garantizando que cada racimo fuese identificado por la predicción con mayor confianza.

Todos los experimentos fueron gestionados mediante scripts Python versionados en GitHub (repositorio interno del proyecto). Cada ejecución registró el hash de commit, la configuración YAML empleada y los pesos entrenados. Las dependencias se almacenaron en requirements.txt y los resultados finales se documentaron en archivos de configuración. yaml y .csv, garantizando la reproducibilidad completa del estudio.

Durante la validación del sistema se analizó su desempeño bajo diferentes condiciones reales de captura, lo cual permitió identificar los factores con mayor impacto en la precisión del modelo. Las variaciones de iluminación mostraron efectos relevantes: la luz difusa generó resultados más estables y precisos, mientras que la iluminación intensa produjo reflejos que dificultaron la detección correcta de algunos racimos. La variación del ángulo de captura también influyó significativamente. Las imágenes cenitales generaron las predicciones más consistentes, seguidas por las laterales, mientras que las oblicuas presentaron mayor complejidad debido a la superposición de estructuras y cambios en la profundidad aparente. Asimismo, la presencia de oclusiones parciales ocasionadas por hojas o tallos disminuyó la capacidad del modelo para identificar correctamente la clase Granel, debido a la pérdida parcial de información relevante.

A partir del análisis de errores se identificó que la clase Granel requería un tratamiento más específico debido a su variabilidad morfológica y a la influencia de oclusiones y contrastes irregulares. Para mejorar su desempeño, se planifica la implementación de minería de errores mediante la recopilación de todos los falsos negativos y falsos positivos generados durante la validación, con el fin de crear un subconjunto especializado que permita refinar el entrenamiento. Adicionalmente, se contempla el uso de umbrales de decisión diferenciados por clase, reduciendo ligeramente el valor asignado a Granel para incrementar su recall sin comprometer excesivamente la precisión general. También se proyecta la incorporación de técnicas de aprendizaje sensible al costo, ajustando los pesos de la función de pérdida para dar mayor relevancia a los errores cometidos en esta clase específica.

Tras el desarrollo del sistema, es necesario avanzar hacia etapas que permitan consolidar su aplicación en escenarios reales. El primer paso consiste en realizar una recolección dirigida de nuevas imágenes enfocadas en casos que el modelo actual presenta como más desafiantes, especialmente aquellos asociados a oclusiones, iluminación extrema y racimos con

variaciones morfológicas poco representadas. Este refuerzo permitirá mejorar la capacidad del modelo para generalizar ante condiciones diversas.

Posteriormente, se plantea llevar a cabo un piloto en un entorno semi-controlado que emule una línea de selección. Este piloto permitirá evaluar el rendimiento del sistema en tiempo real, considerando estabilidad de detección, velocidad de procesamiento y comportamiento ante el movimiento continuo del producto. Asimismo, esta etapa permitirá definir los requerimientos mínimos de hardware e identificar ajustes necesarios para una futura implementación operativa.

Finalmente, será fundamental establecer un mecanismo de monitoreo de deriva del modelo a lo largo del tiempo. La apariencia de las uvas puede variar entre campañas o lotes, lo que puede afectar el desempeño. Un seguimiento periódico permitirá identificar posibles disminuciones en la precisión y activar procesos de actualización mediante reentrenamiento o ajustes de parámetros. Con este enfoque, el sistema no solo podrá demostrar su efectividad inicial, sino también mantenerse consistente y adaptable en usos prolongados.

III. RESULTADOS

El modelo YOLOv11n fue entrenado con un total de 5,137 imágenes, generadas tras aplicar técnicas de aumento de datos sobre un conjunto inicial de 2,125 imágenes capturadas en campo. La distribución del dataset fue de 88% para entrenamiento, 8% para validación y 4% para prueba, lo cual permitió una evaluación precisa del rendimiento general del modelo.

Durante el proceso de entrenamiento, el modelo mostró una convergencia progresiva de las funciones de pérdida, lo que

indica una adecuada capacidad de aprendizaje. En la Figura 9 se presentan las curvas de pérdida de clasificación, objeción e IoU a lo largo de las épocas, evidenciando una estabilización al final del proceso.

El resumen cuantitativo del desempeño del modelo para cada categoría se detalla en la Tabla 4.

Tabla 4
Métricas de desempeño del modelo YOLOv11n por categoría.

Clase	Todas	Granel	Premium
Imágenes	420	306	137
Instancias	1068	923	145
Precisión (P)	0.826	0.761	0.891
Recall (R)	0.748	0.614	0.883
mAP 50	0.837	0.748	0.925
mAP 50:95	0.545	0.404	0.685
F1-Score	0.785	0.680	0.887
Macro-F1	0.783	-	-
Balanced Accuracy	0.748	-	-

Los resultados indican que el modelo obtuvo una alta precisión para la clase “Premium”, alcanzando 89.1% de precisión, 88.3% de recall, y un mAP@50 de 92.5%, lo cual valida su efectividad para identificar racimos de alta calidad.

El modelo identificó uvas Premium con 89.1% de precisión y 88.3% de exhaustividad, superando significativamente el rendimiento en uvas Granel (76.1% precisión, 61.4% exhaustividad). Para validar la fiabilidad estadística de estas métricas y estimar el rango en el que se encuentra el verdadero rendimiento del modelo poblacional, se calcularon los Intervalos de Confianza al 95% (IC95%). Dado que las métricas de precisión y exhaustividad representan proporciones derivadas de una muestra de validación finita, el cálculo se basó en la aproximación normal para proporciones binomiales

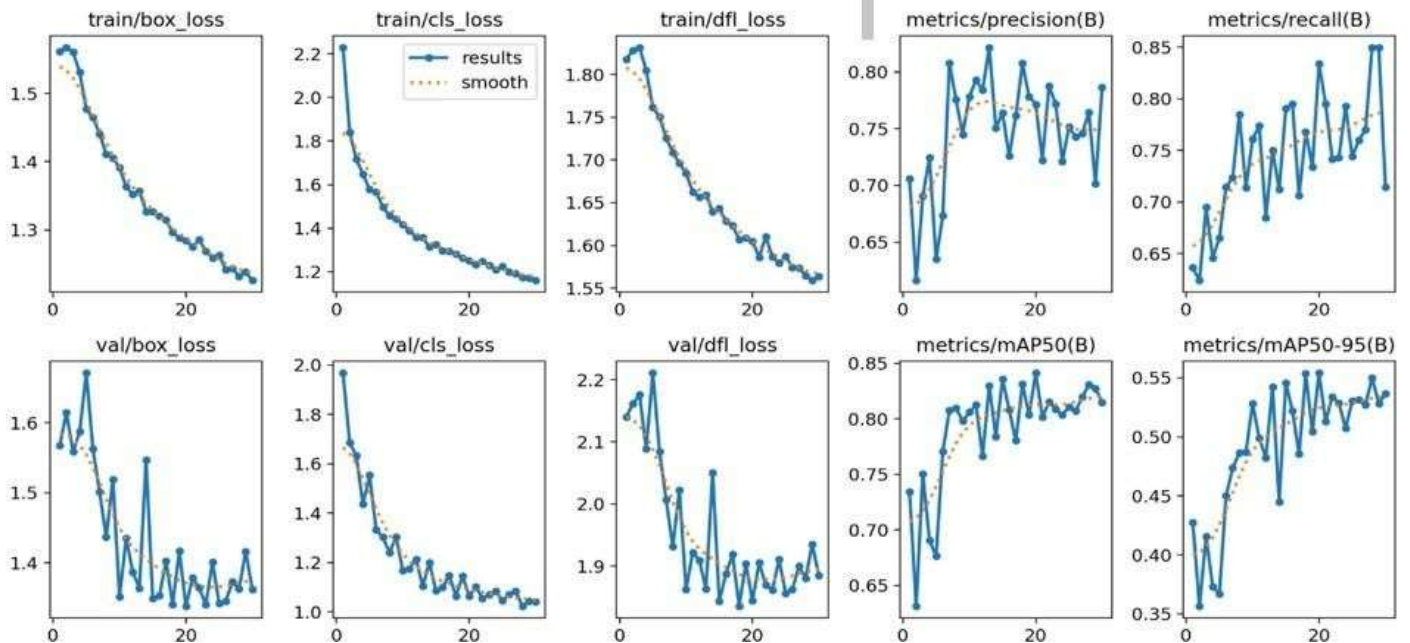


Fig. 9 Curva de pérdidas durante el entrenamiento.

(Intervalo de Wald). La fórmula empleada para determinar los límites superior e inferior de cada métrica ($\hat{p}$) es:

$$IC_{95\%} = \hat{p} \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \quad (5)$$

Donde $\hat{p}$ representa la métrica observada (precisión, recall o mAP), n es el número total de instancias evaluadas en la clase correspondiente (145 para Premium y 923 para Granel) y $Z_{\alpha/2}$ es el valor crítico de la distribución normal estándar para un nivel de confianza del 95% (1.96).

Bajo este análisis, se obtuvo una alta robustez para la categoría Premium, con un intervalo para la precisión situado entre 87.4% y 90.8%, lo que confirma la estabilidad del sistema ante racimos de alta calidad. Por el contrario, la clase Granel presentó intervalos más amplios, específicamente en la métrica de recall (IC95%: 58.0%–64.8%); matemáticamente, esto se debe a que la varianza del término $\hat{p}(1 - \hat{p})$ es mayor cuando la proporción se acerca a 0.5 (50%), cuantificando así la mayor incertidumbre asociada a la heterogeneidad de los defectos visuales en esta categoría superando significativamente el rendimiento en uvas Granel (Fig. 10).

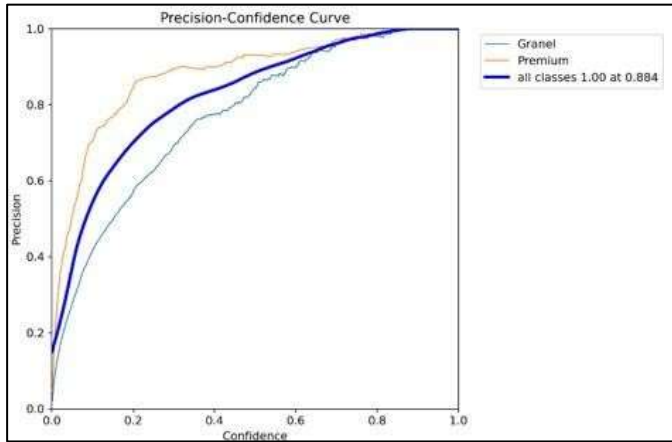


Fig. 10 Curva de precisión-confianza.

La curva F1-Confianza estableció el umbral óptimo en 0.276, donde se maximiza el balance precisión-exhaustividad (F1=0.80). Para Premium se mantiene >90% precisión con umbrales ≥ 0.4 , mientras Granel requiere sensibilidad ajustada.

La métrica mAP@50:95 (54.5%) validó la capacidad del modelo para mantener precisión con umbrales IoU variables (0.5-0.95). La curva Precisión-Exhaustividad confirmó alta confiabilidad en Premium (AP=95.2%) frente a desafíos en Granel (AP=73.1%). La matriz de confusión normalizada (Figura 11) reveló baja confusión interclase (<1%) pero identificó 22% de falsos negativos en Granel, principalmente por confusión con material de fondo (hojas/tallos).

Estos resultados validan la viabilidad técnica para implementación industrial, destacando la necesidad de

optimizar el balance de datos para la clase Granel en futuras iteraciones.

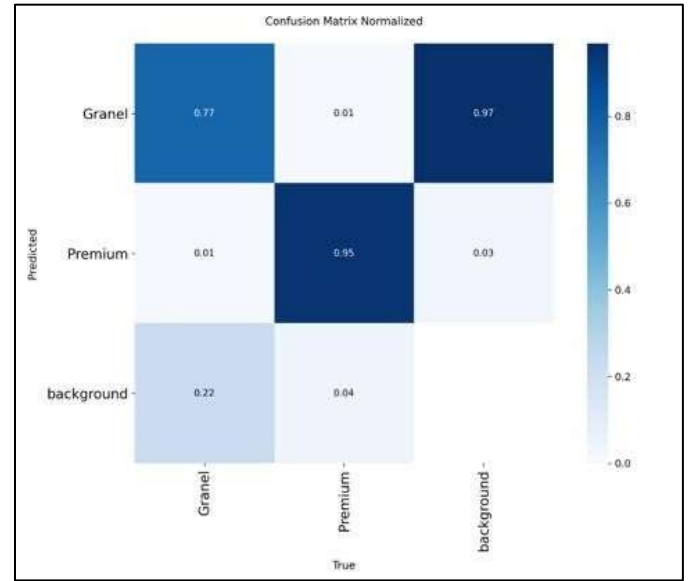


Fig. 11 Matriz de confusión normalizada.

En términos de eficiencia computacional y viabilidad operativa, el modelo YOLOv11n desplegado sobre la GPU NVIDIA RTX 4060 Ti logró una velocidad de procesamiento promedio de 161 cuadros por segundo (FPS). Esto equivale a una latencia total de 6.2 milisegundos por imagen (incluyendo preprocesamiento, inferencia y supresión de no máximos). Este rendimiento supera holgadamente el umbral estándar de 30 FPS requerido para sistemas de visión en tiempo real, confirmando que la arquitectura propuesta es plenamente capaz de operar sincronizada con la velocidad de las cintas transportadoras en una línea de selección continua, garantizando un control de calidad industrial sin generar cuellos de botella computacionales.

IV. DISCUSIÓN

Los resultados cuantitativos alcanzados en esta investigación superan significativamente las métricas reportadas en estudios previos que emplearon arquitecturas más complejas para tareas similares [15], evaluaron el uso de modelos robustos como YOLOv7 para la detección de racimos de uva, reportando un Mean Average Precision (mAP) del 77% y un F1-score del 94%; en contraste, nuestra implementación basada en el modelo compacto YOLOv11n logró un mAP@50 superior del 92.5% para la categoría Premium y una precisión promedio del 89.1%. Esta comparación evidencia que, a pesar de utilizar un modelo con una densidad de parámetros considerablemente menor a la propuesta por [15], se logró una mayor exactitud en la discriminación de clases, validando la eficacia de nuestro proceso de curación y aumento de datos frente a enfoques que priorizan la fuerza bruta computacional.

El análisis del desbalance de clases revela que la diferencia en el recall entre la categoría Premium (88.3%) y Granel (61.4%) se debe a la mayor variabilidad morfológica y la presencia de oclusiones por material orgánico en esta última. Al integrar métricas sensibles al desbalance, se obtuvo una Exactitud Balanceada (Balanced Accuracy) de 74.8% y un Macro-F1 de 78.3%. Estos valores confirman que, aunque el sistema es altamente robusto para la identificación de productos de alta calidad, el sesgo hacia la clase mayoritaria (Premium) limita la sensibilidad en defectos complejos de la clase Granel, validando la necesidad de futuras estrategias de minería de errores o funciones de pérdida ponderadas.

Desde una perspectiva teórica, este estudio expande los hallazgos de [16], quienes introdujeron la arquitectura YOLOv11 destacando la eficiencia de los módulos C3k2 y los mecanismos de atención C2PSA para la detección de objetos generales. Nuestro aporte teórico consiste en validar la capacidad de generalización de estos módulos específicos en un dominio biológico no estructurado, demostrando que la arquitectura ligera no solo optimiza el cómputo, como proponen Wang et al., sino que es capaz de modelar variaciones morfológicas complejas (textura y defectos en uvas) con mayor robustez que sus predecesores. Esto refuta la noción teórica de que se requieren redes profundas y pesadas para capturar detalles finos en entornos agrícolas heterogéneos. En cuanto al aporte práctico, la investigación operacionaliza las proyecciones de [17], quienes en su revisión sistemática identificaron el potencial de la visión artificial para reducir la mano de obra, pero señalaron la falta de implementaciones escalables en entornos reales, nuestro trabajo transforma este potencial en una herramienta de control de calidad tangible (clasificación Premium vs. Granel) lista para su despliegue en planta. El aporte práctico radica en la creación de un pipeline reproducible de bajo costo que cierra la brecha entre la validación académica y la necesidad industrial de sistemas de monitoreo en tiempo real, superando las barreras de adopción tecnológica identificadas en la literatura.

Tabla 5
Comparación cuantitativa con estudios previos en clasificación de uvas

Estudio	Arquitectura	mAP@50	Precisión	Recall	F1-Score
Pinheiro et al. (2023)	YOLOv7	77.0%	98.0%	90.0%	94.0%
Estudio propio (Premium)	YOLOv11n	92.5%	89.1%	88.3%	88.7%
Estudio propio (Granel)	YOLOv11n	74.8%	76.1%	61.4%	68.0%

V. CONCLUSIONES

La presente investigación logró implementar exitosamente un sistema de visión artificial basado en YOLOv11n, cuyos resultados no solo validan su viabilidad técnica, sino que superan los estándares de desempeño reportados recientemente para modelos de mayor complejidad computacional. Al alcanzar un mAP@50 del 92.5% y una precisión del 89.1% en

la categoría Premium, se demostró que es posible obtener una detección de calidad superior sin depender de arquitecturas pesadas, superando las métricas de referencia establecidas por estudios previos en tareas de detección de racimos.

Desde una perspectiva teórica, el estudio aporta evidencia empírica que desafía la noción de que se requieren redes neuronales profundas y masivas para modelar variaciones morfológicas complejas en entornos agrícolas. Se confirmó que las optimizaciones arquitectónicas de YOLOv11n son suficientes para capturar detalles finos como textura y homogeneidad, validando la eficacia de modelos ligeros para tareas de alta precisión biológica. Esto establece un nuevo precedente para el diseño de sistemas de visión en agricultura, donde la eficiencia del modelo no implica un sacrificio en la exactitud. En el ámbito práctico, la investigación logró cerrar la brecha entre la validación académica y la aplicación industrial. A diferencia de enfoques generalistas limitados a la detección de presencia, este sistema se consolidó como una herramienta funcional de control de calidad (clasificación Premium vs. Granel) lista para su despliegue. La arquitectura ligera permite su implementación en hardware de bajo costo, ofreciendo a la industria vitivinícola una solución escalable y objetiva para estandarizar sus procesos de selección, reduciendo la dependencia del juicio manual.

No obstante, se identificaron limitaciones en la clasificación de la categoría Granel (recall del 61.4%), atribuibles a la alta heterogeneidad de sus defectos visuales. Por tanto, se define como trabajo futuro prioritario la aplicación de técnicas de balanceo de clases y minería de errores específicos para esta categoría. Asimismo, se sugiere la validación cruzada y la exploración de despliegues en dispositivos embebidos en el borde (*edge computing*) para consolidar el monitoreo en tiempo real en líneas de producción dinámicas.

REFERENCIAS

- [1] Canizo, Brenda Vanina. (2020). Análisis multielemental y quimiométrico clasificatorio de uvas y productos derivados para evaluar su origen geográfico y trazabilidad. <https://repositorio.unne.edu.ar/handle/123456789/27742>
- [2] Abad Miranda, Jimmy Hendrix; Romero Pérez, Fernando André. (2025). Diseño de un sistema de maceración de uvas automatizado para la elaboración de vino. <https://dspace.ups.edu.ec/handle/123456789/30435>
- [3] Muñoz Castro, Julio Alejandro. (2023). Sistema de control de gestión para una empresa exportadora de fruta fresca. <https://repositorio.uchile.cl/handle/2250/196710>
- [4] Jínez Tapia, José Luis; Acán Carpintero, Jeison Wilmer; Asadobay Paca, Rosana Yadira. (2025). Sistema automático de fumigación de cultivos mediante dron, basado en textura y color dentro de la agricultura 5.0 en un entorno controlado a escala 1:5. <http://dspace.unach.edu.ec/handle/51000/14697>
- [5] Aldana Rodríguez, Didier; Lozano Tafur, Cristian; Orduy Rodríguez, Jaime Enrique. (2024). Uso de redes neuronales convolucionales en UAVs para la detección de anomalías estructurales. <http://repositorio.libertadores.edu.co/items/b11ed420-eca5-491e-9813-d7b688ac0715>
- [6] Forero, Brayan Leonardo. (2024). Modelo para Predicción del Rendimiento de Cultivos de Maíz en Colombia Empleando Aprendizaje Profundo. <https://www.proquest.com/openview/c4978a470731f80fb7e2b3f0c7424c/b1?pq-origsite=gscholar&cbl=2026366&diss=y>

- [7] Ayala Valencia, Felipe. (2024). Prototipo de sistema embebido para la detección de obstáculos mediante redes convolucionales profundas para su integración en un velero autónomo. <https://bibliotecadigital.udea.edu.co/server/api/core/bitstreams/950303c9-8242-4d03-b439-b67e6223fc3a/content>
- [8] Jiménez Gómez, Xabier. (2024). Detección y análisis de datos sobre especies exóticas en biomas diferenciados en apoyo a la biodiversidad empleando la herramienta YOLO y microcontroladores. <https://ruc.udc.es/entities/publication/2fd62404-e9ba-43cf-9756-100d249af86a>
- [9] Revuelta Burgos, Álvaro. (2025). Evaluación, optimización e implementación de modelos convolucionales de visión artificial en dispositivos con recursos limitados. <https://repositorio.unican.es/xmlui/handle/10902/37214>
- [10] Alvarenga Manzanares, Imer Adonis. (2025). Sistema de inteligencia ambiental aplicado a la producción agrícola utilizando SBCs y computación límite. <https://repositorio.upct.es/entities/publication/c6531af0-6759-4df1-8b17-4ac5e56b7725>
- [11] Codes Alcaraz, Ana María. (2025). Teledetección con vehículos aéreos no tripulados y técnicas de aprendizaje profundo para la gestión agrícola. Aplicación a dos casos de estudio en la provincia de Alicante. <https://dspace.umh.es/handle/11000/36487>
- [12] Alejo Orrasco, Jesús. (2023). La implantación de la industria 4.0 en la viticultura: estudio de caso dentro de una bodega en la Ribera del Duero. <https://uvadoc.uva.es/handle/10324/62934>
- [13] Saiz Álvaro, Fátima Aurora. (2018). Estudio de arquitecturas para la extracción y explotación de datos de defectos superficiales mediante técnicas de Deep Learning. <https://reunir.unir.net/handle/123456789/7487>
- [14] García López, Carlos. 2025. Data Augmentation basado en cambios de iluminación usando redes GAN para aplicaciones de reconocimiento de entornos. <https://dspace.umh.es/handle/11000/36042>
- [15] Pinheiro, I., Moreira, G., Queirós da Silva, D., Magalhães, S., Valente, A., Moura Oliveira, P., Cunha, M. y Santos, F. (2023). Solución de aprendizaje profundo basada en YOLO para la detección de racimos de uva y evaluación de lesiones biofísicas. *Agronomía*, 13 (4), 1120. <https://doi.org/10.3390/agronomy13041120>
- [16] Wang, J., Wang, H., Li, Z., Liu, K., Chen, Y., & Xu, X. (2024). YOLOv11: Enhancing Real-Time Object Detection with C3k2, SPPF, and C2PSA Modules. arXiv preprint arXiv:2404.04357. <https://arxiv.org/abs/2404>.
- [17] [C. M. Badgujar, A. Poulouse, and H. Gan, "Agricultural Object Detection with YOLO: A bibliometric and systematic review," arXiv preprint, arXiv:2401.10379, 2024. [Online]. Available: <https://arxiv.org/abs/2401.10379>